GENERAL TO SPECIFIC MODELLING OF EXCHANGE RATE VOLATILITY: A FORECAST EVALUATION[*]

Luc BAUWENS[1] and Genaro SUCARRAT[2]

February 2006

Abstract

The general-to-specific (GETS) approach to modelling is widely employed in the modelling of economic series, but less so in financial volatility modelling due to computational complexity when many explanatory variables are involved. This study proposes a simple way of avoiding this problem and undertakes an out-of-sample forecast evaluation of the methodology applied to the modelling of weekly exchange rate volatility. Our findings suggest that GETS specifications are especially valuable in conditional forecasting, since the specification that employs actual values on the uncertain information performs particularly well.

JEL Classification: C53, F31
Keywords: Exchange Rate Volatility, General to Specific, Forecasting

[1]CORE and Department of Economics, Université catolique de Louvain (Belgium). Email: bauwens@core.ucl.ac.be.

[2]Corresponding author. Department of Economics, Universidad Carlos III de Madrid (Spain), CORE and Department of Economics, Université catolique de Louvain (Belgium). Email: sucarrat@core.ucl.ac.be. Homepage: Http://www.core.ucl.ac.be/~sucarrat/index.html.

1 Introduction

Exchange rate volatility is an issue of great importance for both businesses and policy-makers alike. Hence businesses use volatility models as tools in their risk management and as input in derivative pricing, whereas policymakers use them to acquire knowledge about what and how economic factors impact upon exchange rate volatility for informed policymaking. Most volatility models are highly non-linear and thus require complex optimisation algorithms in empirical application. For models with few parameters and few explanatory variables this may not pose unsurmountable problems. But as the number of parameters and explanatory variables increases the resources needed for reliable estimation and model validation multiply. Indeed, this may even become an obstacle to the application of certain econometric modelling strategies, as for example argued by McAleer (2005) regarding automated general-to-specific (GETS) modelling of financial volatility.[1] GETS modelling is particularly suited for explanatory econometric modelling since it provides a systematic framework for statistical economic hypothesis-testing, model development and model (re-)evaluation, and the methodology is relatively popular among large scale econometric model developers and proprietors. However, since the initial model formulation typically entails many explanatory variables this poses challenges already at the outset for computationally complex models.

In this study we overcome the computational challenges traditionally associated with the application of the GETS methodology in the modelling of financial volatility by modelling a measure of observed volatility (squared return) directly within a single equation exponential model of observable volatility (EMOV) framework with ordinary least squares (OLS) estimation. This enables us to apply GETS on a general specification with, in our case, a constant and twenty four regressors, including lags of log of volatility, an asymmetry term, a skewness term, seasonality variables, and economic covariates. Compared with models of the autoregressive conditional heteroscedasticity (ARCH) and stochastic volatility (SV) classes we estimate and simplify our specification effortlessly, and obtain a parsimonious encompassing specification with uncorrelated homoscedastic residuals and relatively stable parameters. Moreover, our out-of-sample forecast evaluation suggests that GETS specifications are especially valuable in conditional forecasting, since the specification that employs actual values on the uncertain information performs particularly well.

The rest of the paper is divided into four sections. In the next section we justify the methodology of our study. Then we present the models in section 3, whereas section 4 contains the results of the out-of-sample forecast exercise. Finally we conclude in section 5.

[1]GETS modelling is also sometimes referred to as the "LSE methodology", after the institution in which the methodology to a large extent originated in, and sometimes even "British econometrics", see Gilbert (1989), Mizon (1995) and Hendry (2003).

2   Explanatory exchange rate volatility modelling

This section justifies the research design of our study and presents the EMOV in more detail. In the first subsection we give some considerations regarding the measurement of volatility forecast accuracy in an explanatory context. Then, in the second subsection, we give a brief overview of the GETS methodology. Finally, in the third subsection we describe the EMOV and compare it with the more common ARCH and SV families of models.

2.1   The explanatory perspective

According to the Merriam-Webster Online dictionary the etymological origin of the word "volatility" is the latin *volatilis*, which is a derivative of *volare*, "to fly". Although one of the meanings of volatility still is "flying" or "having the power to fly", the term typically carries a rather different and specific meaning in financial econometrics, namely the conditional standard deviation (or variance) of a financial return. Volatility is thus defined as a latent or unobservable variable, deterministic or stochastic. In this study, however, we find it useful to make a distinction between observed and unobserved—that is, latent—volatility, where absolute and squared returns are examples of observable definitions of volatility.

A view that has gained widespread acceptance lately is that volatility forecasts should be compared with realised volatility, that is, sums of squared intra-period returns, rather than (say) squared or absolute period returns, see Andersen and Bollerslev (1998), and Andersen et al. (1999). Because although "squared returns constitute an unbiased estimator for the latent volatility factor, they also embody a large idiosyncratic component that is unrelated to the actual volatility driving the market over the observation interval" (Andersen and Bollerslev 1999, p. 458). In other words, the "large unrelated component" tends to contaminate volatility forecast comparisons leading to possibly erroneous conclusions. As a remedy they suggest that volatility forecasts are evaluated against realised volatility, which they define as the sum of squared intra period squared returns, since "it corresponds directly to the notion of volatility entertained in diffusion models..." (Andersen et al. 1999, p. 458).

There are at least three objections to this view from an explanatory perspective. The first is that one is restricted to work within the continuous time diffusion framework as if it constituted *the* "true" model, something that is particularly inappropriate in explanatory econometric modelling. Economic events have temporal extension and stand at the end of chains of economic events, each with temporal extension. In other words, it takes time for one event to bring about another. So as the time increment goes to zero, so does the explanatory potential of explanatory modelling. Second, absolute and squared returns are measures of the total period variation in the exchange rate, whereas realised volatility is an estimate of a latent variable. As a consequence, restricting one's focus on latent volatility means that one disregards how the latent variable transmits to observable volatility, the magnitude which many decision makers ultimately care about and define their loss function in terms of. Andersen and Bollerslev (1998, p. 890) have argued that rational "financial

decision making hinges on the anticipated future volatility and not the subsequent [observed] squared returns" (p. 890). We agree that expectations play a part, but we do not agree in the latter part of the sentence, namely that observed volatility does not matter. Rational financial decision making also depends on the anticipated possible consequences of the discrepancy between forecasts and actual values, and an important part of rational financial decision making consists of continuously evaluating to what extent forecasts differ from actual values and how this difference can be reduced. This is not an argument against latent volatility approaches as such, only an emphasis of the fact that different questions call for different definitions of volatility. As a contrast, if our aim were to price a derivative having decided already at the outset to do it by means of a continuous time diffusion model, then the latent definition of volatility (quadratic variation) in the diffusion model would be more appropriate. Third and finally, in explanatory econometric volatility modelling the error term plays an informative role since it is an objective measure—in the sense that it does not rely upon any assumptions about what the "true" model is—of how successful or unsuccessful the specification is compared with others. Indeed, even in derivative pricing one way of evaluating the precision of the underlying process that is assumed to generate the price series is by comparing its forecasts with actual values. For these reasons our focus is on the discrepancy between forecasts and actual squared returns rather than (say) realised volatility as defined by Andersen and Bollerslev (1998).

2.2   GETS modelling

A fundamental cornerstone of the GETS methodology is that empirical models are derived, simplified representations of the complex human interactions that generate the data. So instead of postulating a uniquely "true" model or paradigm, the aim is to develop "congruent" models within the statistical framework of choice. The exact definition of congruency is given below, but in brief a congruent model is a theory informed specification that is data-compatible and which constitutes a "history-repeats-itself" representation (stable parameters, innovation errors).[2]

In econometric practice GETS modelling proceeds in cycles of three steps: 1) Formulate a general unrestricted model (GUM) which is congruent, 2) simplify the model sequentially in an attempt to derive a parsimonious congruent model while at each step checking that the model remains congruent, and 3) test the resulting congruent model against the GUM. The test of the final model against the GUM serves as a parsimonious encompassing test, that is, a test of whether important information is lost or not in the simplification process. If the final model is not congruent or if it does not parsimoniously encompass the GUM, then the cycle starts all over again by re-specifying the GUM. As such the GETS methodology treats modelling as a process, where the aim is to derive a parsimonious congruent encompassing model while at the same time acknowledging that "the currently best available model" (Hendry and Richard 1990, p. 323) can always be improved.

---

[2]The term "congruent" is borrowed from geometry: By "analogy with one triangle which matches another in all respects, the model matches the evidence in all measured respects." (Hendry 1995, p. 365)

GETS modelling derives its basis from statistical reduction theory in general and Hendry's reduction theory (1995, chapter 9) in particular,[3] which is a probabilistic framework for the analysis and classification of the simplification errors associated with empirical models. The "theory offers", in Hendry's own words, "an explanation for the origin of all empirical models" (1997, p. 174) in terms of twelve "reduction operations conducted implicitly on the DGP..." (1995, p. 344),[4] and GETS modelling seeks to mimic reduction analysis by evaluating at each reduction whether important information is lost or not. Evaluation of any empirical model can take place against six types of information-sets, namely 1) past data, 2) present data, 3) future data, 4) theory information, 5) measurement information and 6) rival models, and with each of these types we may delineate an associated set of properties that a model should exhibit in order to be considered as a satisfactory, simplified representation of the DGP:[5]

1. Innovation errors. For a model to be a satisfactory representation of the process that generated the data, what remains unexplained should vary unsystematically, that is, the errors should be innovations. In practice this entails checking whether the residuals are uncorrelated and homoscedastic.

2. Weak exogeneity. This criterion entails that conditioning variables are weakly exogenous for the parameters of interest.

3. Constant, invariant parameters of interest. Models without stable parameters are unlikely to be successful forecasting models, so this is a natural criterion if successful forecasting is desirable.

4. Theory consistent, identifiable structures. To ensure that a model has a basis in economic reality it should be founded in economic argument.

5. Data admissibility. In the current context, an example of a volatility model that violates this criterion is one that produces negative volatility forecasts.

6. Encompassing of rival models. A model encompasses another if it accounts for its results. Within the three-step cycle of GETS modelling sketched above, a parsimonious encompassing test is undertaken when the final model is tested against the GUM. If no or sufficiently little information is lost then the final model accounts for the results of the GUM.

Models characterised by the first five criteria are said to be congruent, whereas models that also satisfy the sixth are said to be encompassing congruent.

It is important to distinguish between two aspects of the GETS methodology, namely the properties a model (ideally) should exhibit on the one hand, that is, congruent encompassing, and the process of deriving it on the other, that is, general-to-specific search.

---

[3]Other expositions of the GETS methodology and its foundations are Hendry and Richard (1990), Gilbert (1990), Mizon (1995) and Jansen (2002).

[4]DGP stands for data-generating process.

[5]See Hendry (1995, pp. 362-367) and Mizon (1995) for further discussion .

Contrary to what the name of the GETS methodology may suggest it is actually the former that is of greatest importance. In the words of Hendry, "the credibility of the model is not dependent on its mode of discovery but on how well it survives later evaluation of all of its properties and implications..." (1987, p. 37). However, there is no secret that general-to-specific search for the "currently best available" specification is the preferred approach by the proponents of the GETS methodology. In addition to the fact that it mimics reduction analysis at least four additional important reasons can be listed:[6] The search for the currently best available specification is ordered since any specification obtained in the search is nested within the GUM; in statistical frameworks where adding regressors reduces the residual variance—as for example in the linear model with OLS estimation—the power in hypothesis testing increases; the GETS methodology provides a systematic approach to economic hypothesis testing; and finally compared with unsystematic searches GETS search is resource efficient, see Hendry and Krolzig (2004).

2.3   Models of exchange rate volatility

If s_t denotes the log of an exchange rate and r_t the log-return, that is, $r_t = \Delta s_t = s_t - s_{t-1}$, then we will refer to r_t^2 as observed volatility. The EMOV is given by

$$r_t^2 = \exp(\mathbf{b}'\mathbf{x}_t + u_t), \tag{1}$$

where $\mathbf{b}$ is a parameter vector, $\mathbf{x}_t$ is the vector of conditioning variables and $\{u_t\}$ is a sequence of mutually uncorrelated and homoscedastic variables each with conditional mean equal to zero. The exponential specification (1) is motivated by several reasons. The most straightforward is that it results in simpler estimation, in particular when many explanatory variables are involved. Under the assumption that $\{r_t^2 = 0\}$ is an event with probability zero, then consistent and asymptotically normal estimates of $\mathbf{b}$ can be obtained almost surely with OLS under standard assumptions, since

$$\log r_t^2 = \mathbf{b}'\mathbf{x}_t + u_t \quad \text{with probability 1.} \tag{2}$$

Another motivation for the exponential specification, which was first pointed to by Geweke (1986) and Pantula (1986), and which subsequentially led Nelson (1991) to formulate the exponential general ARCH (EGARCH) model, is that it ensures positivity. This is particularly useful in empirical analysis because it ensures that fitted values of volatility are not negative. Finally, another attractive feature of the exponential specification is that it produces residuals closer to the normal in (2) and thus presumably leads to faster convergence of the OLS estimator. In other words, the log-transformation is likely to result in sounder inference regarding $\mathbf{b}$ in (2) when an asymptotic approximation is used. Applying the conditional expectation operator on observed volatility in (1) gives

$$E(r_t^2|\mathcal{I}_t) = \exp(\mathbf{b}'\mathbf{x}_t) \cdot E[\exp(u_t)|\mathcal{I}_t], \tag{3}$$

---

[6]See Campos et al. (2005) for a more complete discussion.

where $\mathcal{I}_t$ denotes the information set in question. An estimate of conditional observed volatility is readily obtained if either $\{u_t\}$ is IID or if the $\{\exp(u_t)\}$ are uncorrelated, since the formula $\frac{1}{T}\sum_{t=1}^{T}\exp(\hat{u}_t)$ then provides a consistent estimate of the normalising constant $E[\exp(u_t)|\mathcal{I}_t]$.

To see the relation between the EMOV and the ARCH and SV families of models, recall that the latter two decompose log-returns into a conditional mean μ_t and a remainder e_t

$$r_t = \mu_t + e_t, \tag{4}$$

where $e_t = \sigma_t z_t$ and $\{z_t\}$ is a sequence of mutually uncorrelated random variables with conditional mean equal to zero and conditional variance equal to one. This implies

$$Var(r_t|\mathcal{I}_t) = E(\sigma_t^2 z_t^2|\mathcal{I}_t). \tag{5}$$

If σ_t^2 follows a non-stochastic autoregressive process, then (4) belongs to the ARCH family and the conditional variance in (5) reduces to σ_t^2. A common example is the GARCH(1,1) of Bollerslev (1986) where

$$\sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2. \tag{6}$$

If σ_t^2 on the other hand follows a stochastic autoregressive process, then (4) belongs to the SV family of models. In the special case where σ_t and z_t are independent the conditional variance equals $E(\sigma_t^2|\mathcal{I}_t)$.

The relation between EMOV and the ARCH and SV families can be viewed in at least three ways. The first is to treat the EMOV as a non-nested alternative to the ARCH and SV families. At first sight one might object that the EMOV does not account for the influence of variables in the conditional mean of returns, but this is incorrect. Variables normally included in the conditional mean equation can appear in $\mathbf{x}_t$. For example, r_{t-1} is often included as a regressor in the conditional mean equation of ARCH models of exchange rate returns, and one way to account for its potential influence in the EMOV is by including $\log r_{t-1}^2$ in $\mathbf{x}_t$. A second way of viewing the EMOV is to interpret it as an approximation to observed volatility in the ARCH and SV families. Recall that expected observed volatility within the ARCH family[7] is

$$E(r_t^2|\mathcal{I}_t) = \mu_t^2 + \sigma_t^2. \tag{7}$$

In words, the total expected exchange rate variation consists of two components, the squared conditional mean μ_t^2 and the conditional variance σ_t^2. As Jorion (1995, footnote 4 p. 510) has noted σ_t^2 typically dwarfs μ_t^2 with a factor of several hundreds to one,[8] so the "de-meaned" approximation

$$\mu_t^2 + \sigma_t^2 \approx \sigma_t^2 \tag{8}$$

---

[7] No generality is lost by only considering the ARCH family since the same type of argument applies with respect to the SV family under standard assumptions.

[8] Jorion noted that, on daily data, the factor is typically 700 to 1. In our case the median of the fitted values of σ_t^2 is between 500 and 600 times greater than the median of the fitted values of μ_t^2 in the GARCH(1,1) and EGARCH(1,1) specifications of subsection 3.2.

is often reasonably good in practice. The third way to view the relation between EMOV and the ARCH and SV families of models requires that errors are interpreted as designed, and enables us to model EMOV and ARCH/SV specifications jointly. For example, consider the bivariate system

$$r_t = \mathbf{a}'\mathbf{z}_t + e_t, \qquad \log r_t^2 = \mathbf{b}'\mathbf{x}_t + u_t, \tag{9}$$

where the errors are interpreted as designed, that is,

$$e_t = r_t - E(r_t|\mathbf{z}_t \cup \mathbf{x}_t), \qquad u_t = \log r_t^2 - E(\log r_t^2|\mathbf{z}_t \cup \mathbf{x}_t). \tag{10}$$

The first equation in (9) with r_t on the left hand side can be formulated as an ARCH or SV specification and the second as an EMOV. This gives rise to testable questions like weak exogeneity (valid conditioning with respect to the parameters of interest), strong exogeneity (whether lags of $\log r_t^2$ predicts r_t and whether lags of r_t predicts $\log r_t^2$), whether there is a presence or absence of ARCH/SV in the residuals of the first equation, and ultimately whether (say) the first equation forecast encompasses the second in predicting r_t^2. For example, suppose we obtain a parsimonious version of (9) which exhibits weak and strong exogeneity, and that a subsequent out-of-sample forecast comparison of r_t^2 reveals that the EMOV outperforms the ARCH/SV model. Then we are (statistically) justified in only focusing on the EMOV in our investigation of r_t^2, since the equation of r_t is not needed for valid inference nor for improved prediction. Alternatively, if the out-of-sample forecast comparison suggests that the ARCH/SV specification is superior in predicting r_t^2, then we would be right in disregarding the EMOV specification.

3   Data and empirical models

This section presents the data of our study and our empirical forecast models, and proceeds in four steps. The first subsection describes our data in brief (the data appendix provides more details) and introduces notation, whereas the next three subsections describe our forecast models. The economic motivation, justification and interpretation of the variables have been dealt with in greater length elsewhere, see Bauwens et al. (2006), so here we concentrate on the statistical properties of the models. The second subsection contains specifications that condition on both "certain" and "uncertain" information. With certain information we mean information that is predictable with a high degree of certainty, for example past values, holidays, etc. With uncertain information we mean information that is not predictable with a high degree of certainty. Typical examples would be contemporaneous values of economic variables, etc. The motivation behind the distinction between certain and uncertain information is that it enables us to gauge the potential forecast precision in the ideal case where the values of the uncertain information are correct. This is of particular interest since the GETS methodology often is championed for its ability to develop models appropriate for scenario analysis (counterfactual analysis, policy analysis, conditional forecasting, etc.), where conditioning on uncertain information plays an important part. The distinction is also of practical interest, since it enables us to investigate

whether GETS models with uncertain information improve upon the forecast accuracy of models without uncertain information, since the uncertain information would have to be forecasted in a realistic forecast setting. The third subsection contains specifications with certain information only, whereas the fourth and final subsection contains the benchmark or "simple" specifications that serve as a point of comparison. These models are relatively parsimonious and require little development and maintenance effort, thus the label "simple", and they have a documented forecasting record. Their motivation is that an important issue is whether GETS derived specifications improve upon the forecast accuracy provided by simple models.

3.1   Data and notation

Our weekly data span the period 8 January 1993 to 25 February 2005, a total of 634 observations, and the details of the data transformations and the data sources are given in the appendix.[9] In order to undertake out-of-sample accuracy evaluation we split the sample in two. The estimation sample is 8 January 1993 - 26 December 2003 (573 observations), and the reason we split the sample at this point is that it then corresponds to that of Bauwens et al. (2006). The remaining 61 observations are used for the out-of-sample analysis. The exchange rate in question is the closing value of the BID NOK/EUR in the last trading day of the week and is denoted by S_t. Note that before 1 January 1999 we use the BID NOK/DEM exchange rate converted to euro-equivalents with the official conversion rate 1.95583 DEM = 1 EURO. The weekly return is given by $r_t = \log S_t - \log S_{t-1}$, and the weekly variance by $V_t^w = r_t^2$. We will make extensive use of the log-transformation applied on volatilities and generally we will follow the convention of denoting such variables in lower case. For example, the log of squared NOK/EUR returns is denoted v_t^w and defined as $v_t^w = \log V_t^w$. Graphs of S_t, r_t and v_t^w are contained in figure 1.

In addition to lags of volatilites we also include several other regressors in our specifications, including a low frequency version of weekly realised variance which we denote V_t^r with its log counterpart as v_t^r, and a weekly range based volatility measure which we denote V_t^{hl} with its log counterpart as v_t^{hl}. The weekly realised variance is constructed by using the opening and closing values of the trading days of the week, whereas the range based measure is constructed by using the minimum and maximum values over the week. To account for the possibility of skewness and asymmetries in r_t we use the lagged return r_{t-1} for the latter, and an impulse dummy ia_t equal to 1 when returns are positive and 0 otherwise for the former. We also include variables intended to account for the impact of holidays and seasonal variation. These are denoted h_{lt} with $l = 1, 2, \ldots, 8$, see the appendix for further details. As a measure of variation in market activity we use the relative change in the number of quotes. More precisely, if we denote the number of quotes

---

[9]Over this period Norway experienced three different types of exchange rate regimes. Loosely, until 1998 the central bank of Norway (Norges Bank) actively sought to stabilise the Norwegian krone against its main trading partners, then it shifted to partial inflation targeting before it was instructed by the Ministry of Finance to fully pursue inflation targeting in March 2001. For more details, see Bauwens et al. (2006).

in week t by Q_t and its log-counterpart by q_t, we use Δq_t as our measure of the relative change in market activity from one week to the next. The variable q_t can be interpreted as the general level of market activity due to (say) the number of traders active or other institutional characteristics. As a measure of general currency market turbulence we use EUR/USD-volatility. If $m_t = \log (\text{EUR/USD})_t$, then Δm_t denotes the weekly return of EUR/USD, M_t^w stands for weekly volatility and m_t^w is its log-counterpart. The petroleum sector plays a major role in the Norwegian economy, so it makes sense to also include a measure of oilprice volatility. If the log of the oilprice is denoted o_t, then the weekly return is Δo_t, weekly volatility is O_t^w with o_t^w as its log-counterpart. We proceed similarly with Norwegian and US stock market variables. If x_t denotes the log of the main index of the Oslo stock exchange, then the associated variables are Δx_t, X_t^w and x_t^w. In the US case u_t is the log of the New York stock exchange (NYSE) index and the associated variables are Δu_t, U_t^w and u_t^w. The foreign interest-rate variables that we include are constructed using an index made up of the short term market interest-rates of the EMU countries. Specifically, if IR_t^{emu} denotes this interest-rate index then we include a variable that is denoted ir_t^{emu} and which is defined as $(\Delta IR_t^{emu})^2$. The Norwegian interest-rate variables that we include are constructed using the main policy interest rate variable of the Norwegian central bank. Let F_t denote the main policy interest rate in percentages and let ΔF_t denote the change from the end of one week to the end of the next. Furthermore, let I_a denote an indicator function equal to 1 in the period 1 January 1999 - Friday 30 March 2001 and 0 otherwise, and let I_b denote an indicator function equal to 1 after 30 March 2001 and 0 before. In the first period the Bank pursued a "partial" inflation targeting policy, whereas in the second it pursued a "full" inflation targeting policy. We then have $\Delta F_t^a = \Delta F_t \times I_a$ and $\Delta F_t^b = \Delta F_t \times I_b$, respectively, and f_t^a and f_t^b stand for $|\Delta F_t^a|$ and $|\Delta F_t^b|$, respectively. Finally, we also include a step dummy sd_t equal to 0 before 1997 and 1 after to account for a structural increase in volatility.

3.2 Models with both certain and uncertain information

This subsection presents our models with both certain and uncertain information. Specifically they are

GUM EMOV1: $\quad v_t^w = b_0 + b_1 v_{t-1}^w + b_2 v_{t-2}^w + b_3 v_{t-3}^w + b_4 v_{t-4}^w + b_5 q_t + b_6 \Delta q_t + b_7 m_t^w$
$$+ b_8 o_t^w + b_9 x_t^w + b_{10} u_t^w + b_{11} f_t^a + b_{12} f_t^b + b_{13} ir_t^{emu}$$

$$+ b_{14} sd_t + b_{15} ia_t + b_{16} r_{t-1} + \sum_{l=1}^{8} b_{16+l} h_{lt} + e_t \quad (11)$$

GETS EMOV1: $\quad v_t^w = b_0 + b_2 (v_{t-2}^w + v_{t-3}^w) + b_6 \Delta q_t + b_9 (x_t^w + u_t^w)$
$$+ b_{12} f_t^b + b_{13} ir_t^{emu} + b_{14} sd_t + e_t, \quad (12)$$

GETS EMOV2: $\quad v_t^w = b_0 + b_2 (v_{t-2}^w + v_{t-3}^w) + b_9 (\bar{x}^w + \bar{u}^w) + b_{13} \bar{ir}_t^{emu} + b_{14} sd_t + e_t, \quad (13)$

GETS EMOV3: $v_t^w = b_0 + b_2(v_{t-2}^w + v_{t-3}^w) + b_9(\dot{x}^w + \dot{u}^w) + b_{13}\dot{ir}_t^{emu} + b_{14}sd_t + e_t,$ (14)

where $\{e_t\}$ is a sequence of innovation errors. The first specification GUM EMOV1 is a general and unrestricted model with both known and unknown information, whereas the second which is labelled GETS EMOV1 is the GETS derived counterpart. Of these two only the second will be used in our out-of-sample study. The second specification is obtained by setting the first as the general unrestricted specification, and then testing restrictions regarding the parameters with Wald-tests before the final specification is tested against the GUM. It should be noted that we only perform a single specification search where, at each step, we remove the regressor with highest p-value. Hoover and Perez (1999) have pointed out that performing only a single simplification search might result in "path dependence", in the sense that a relevant variable being removed early on in the search whereas irrelevant variables that proxy its role are retained. However, the software PcGets version 1.0 (see Hendry and Krolzig 2001), which automates GETS multiple-path simplification search, produces a specification almost identical to (12), the only difference being that v_{t-2}^w is not retained. So path-dependence does not appear to be a problem in our case. This is consistent with White's (1990) theorem, which implies that the path dependence problem reduces as the size of the sample increases.[10] In the generation of GETS EMOV1 forecasts two steps ahead and onwards we use forecasted values of v_t^w and observed values of the other covariates. In other words, the forecasts of GETS EMOV1 are generated as if the unknown conditioning information is known. As such the accuracy of GETS EMOV1 constitutes an indication of its potential for scenario analysis (policy analysis, conditional forecasting, counterfactual analysis, etc.), since its accuracy will reflect its potential of yielding accurate forecasts under the assumption that the unknown information is correct. The third and fourth specifications serve as a contrast to this hypothetical situation and try to mimic more realistic circumstances by using the parameter estimates of GETS EMOV1, and by using simple forecasting rules for the uncertain information. In GETS EMOV2 the variables Δq_t and f_t^b are set equal to zero, and x_t^w, u_t^w and ir_t^{emu} are set equal to their sample averages $\bar{x}^w$, $\bar{u}^w$ and $\bar{ir}_t^{emu}$ over the period 1 January 1999 - 26 December 2003.[11] In other words, variables that would have to be forecasted in a realistic setting are either set to zero or to their presumed, future sample averages. GETS EMOV3 proceeds similarly with a single difference. Instead of averages the medians of x_t^w, u_t^w and ir_t^{emu}, denoted $\dot{x}^w$, $\dot{u}^w$ and $\dot{ir}_t^{emu}$, are used.

Estimation results and recursive parameter stability analysis of the first two specifications are contained in table 1, and in figures 2 and 3. Both GUM EMOV1 and GETS EMOV1 exhibit innovation errors in the sense that the nulls of no serial correlation, no autoregressive conditional heteroscedasticity and no heteroscedasticity are not rejected at the 10% significance level, and the recursive parameter stability analysis suggests parameters are relatively stable. For both GUM EMOV1 and GETS EMOV1 the Chow forecast

---

[10]Our sample of 573 observations is considerably larger than those investigated by Lovell (1983), Hoover and Perez (1999) and Hendry and Krolzig (1999), the sequence of studies that resulted in PcGets. Whereas Lovell (1983) used only 23 observations, the other two studies employed a maximum of 140 observations.

[11]This sample was chosen because the volatility of r_t looks relatively stable over this period.

and breakpoint tests do not signify at the 1% level, but the 1-step forecast tests on the other hand show some signs of instability.[12] The number of spikes that exceeds the 1% critical value in the break-point tests is 11 and 13, respectively. This suggests the presence of some structural instability since on average we would expect only 5 spikes to exceed the 1% critical value (1% of 473 is just below 5). [13]

3.3 Models with certain information

This subsection contains our specifications with known or relatively certain information. Specifically they are

GUM EMOV4:
$$v_t^w = b_0 + b_1 v_{t-1}^w + b_2 v_{t-2}^w + b_3 v_{t-3}^w + b_4 v_{t-4}^w + b_{14} sd_t$$
$$+ b_{16} r_{t-1} + \sum_{l=1}^{8} b_{16+l} h_{lt} + e_t, \quad (15)$$

GETS EMOV4:
$$v_t^w = b_0 + b_2 (v_{t-2}^w + v_{t-3}^w) + b_{14} sd_t + b_{18} h_{2t} + e_t, \quad (16)$$

Realised EMOV5:
$$v_t^w = b_0 + b_1 v_{t-1}^r + b_{14} sd_t + b_{18} h_{2t} + e_t, \quad (17)$$

Range EMOV6:
$$v_t^w = b_0 + b_1 v_{t-1}^{hl} + b_{14} sd_t + b_{18} h_{2t} + e_t, \quad (18)$$

GARCH(1,1)+:
$$r_t = b_0 + b_1 r_{t-1} + e_t, \qquad e_t = \sigma_t z_t,$$
$$\sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma_1 h_{2t}, \quad (19)$$

EGARCH(1,1)+:
$$r_t = b_0 + b_1 r_{t-1} + e_t, \qquad e_t = \sigma_t z_t,$$
$$\log \sigma_t^2 = \omega + \alpha |\frac{e_{t-1}}{\sigma_{t-1}}| + \beta \log \sigma_{t-1}^2 + \gamma_0 \frac{e_{t-1}}{\sigma_{t-1}} + \gamma_1 h_{2t}, \quad (20)$$

where σ_t is the conditional standard deviation of r_t, and $\{z_t\}$ is a sequence of random variables each with mean equal to zero conditional on the information set in question, and each with variance equal to one conditional on the same information set. The first specification GUM EMOV4 is a general formulation nested within GUM EMOV1 but containing only "certain" conditioning information, that is, past and relatively certain contemporaneous information (holiday variables). The second specification GETS EMOV4 is obtained

---

[12]If t denotes the sample size, k the number of parameters in $\mathbf{b}$ and M the observation at which recursive estimation starts, then for $t = M, \ldots, T$ the 1-step, breakpoint and forecast tests are computed in PcGive as $F(1, t - k - 1)$, $F(T - t + 1, t - k - 1)$ and $F(t - M + 1, M - k - 1)$, respectively.

[13]The number 473 is due to the fact that the recursive estimation was initialised at observation number 100.

through GETS-analysis of GUM EMOV4. The third specification Realised EMOV5 uses a low frequency version of lagged realised weekly volatility v_{t-1}^r as predictor together with the structural step dummy sd_t and the Good Friday holiday variable h_{2t}, whereas Range EMOV6 is identical to Realised EMOV5 except that it replaces lagged weekly realised volatility with lagged weekly range volatility. In the fifth and sixth specifications a constant b_0, lagged return r_{t-1} and h_{2t} are added to "plain" GARCH(1,1) and EGARCH(1,1) specifications. In addition to the fact that the conditional variance σ_t^2 is modelled exponentially, the EGARCH differs from the GARCH by the inclusion of an asymmetry term $\frac{e_{t-1}}{\sigma_{t-1}}$ in the conditional variance specification. A value of γ_0 unequal to zero implies asymmetry and $\gamma_0 < 0$ in particular implies leverage, that is, that returns are negatively correlated with last period's volatility. The higher $|\beta|$ the higher persistence, and a necessary condition for covariance stationarity is $|\beta| < 1$, see Nelson (1991).

The estimation results of the six specifications are contained in tables 2, 3 and 4, and recursive parameter stability analysis of GUM EMOV4 and GETS EMOV4 in figures 4 and 5. The first four specifications all exhibit innovation errors in the sense that the nulls of no serial correlation, no autoregressive conditional heteroscedasticity and no heteroscedasticity are not rejected at conventional significance levels, and the recursive parameter stability analysis for GUM EMOV4 and GETS EMOV4 are similar to those of GUM EMOV1 and GETS EMOV1 above. Both GARCH(1,1)+ and EGARCH(1,1)+ exhibit uncorrelated standardised residuals and squared standardised residuals according to the diagnostic tests, and the lagged return r_{t-1} in the mean equation is negative as commonly found for exchange rates, but not significant. The estimates of $\alpha + \beta$ (0.129 + 0.877 = 1.006) and β (0.983) are very close to 1. This is usually interpreted as an indication of a strong persistence of shocks on the conditional variance, but in this case it is due to the structural break around the beginning of 1997, see figure 1. Finally, the value of γ_0 is insignificantly different from zero which suggests that the symmetry imposed by the GARCH model is not restrictive.

3.4 Simple models

Our benchmark or simple models are all ARCH-specifications, and specifically they are

$$\text{Historical:} \qquad r_t = e_t = \sigma_t z_t, \qquad \sigma_t^2 = \omega \tag{21}$$

$$\text{RiskMetrics:} \qquad r_t = e_t = \sigma_t z_t, \qquad \sigma_t^2 = 0.06 e_{t-1}^2 + 0.94 \sigma_{t-1}^2 \tag{22}$$

$$\text{EWMA:} \qquad r_t = e_t = \sigma_t z_t, \qquad \sigma_t^2 = \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{23}$$

$$\text{GARCH(1,1):} \qquad r_t = e_t = \sigma_t z_t, \qquad \sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{24}$$

$$\text{EGARCH(1,1):} \quad r_t = e_t = \sigma_t z_t, \qquad \log \sigma_t^2 = \omega + \alpha \left| \frac{e_{t-1}}{\sigma_{t-1}} \right| + \beta \log \sigma_{t-1}^2 + \gamma \frac{e_{t-1}}{\sigma_{t-1}}, \tag{25}$$

where $\{z_t\}$ is characterised as above. The first specification labelled Historical is a GARCH(0,0) estimated on the sample 1/1/1999 - 26/12/2003 (261 observations). In other words, it is the ARCH-counterpart of the sample variance since it models volatility as non-varying, and the sample was chosen because volatility appears relatively stable graphically over this period. Failure to beat the historical variance is detrimental to models of the ARCH-class, since this essentially undermines their *raison d'être*. The second specification is an exponentially weighted moving average (EWMA) with parameter values suggested by RiskMetrics (Hull 2000, p. 372).[14] RiskMetrics proposed these values after having compared a range of combinations on various financial time series. The third specification is an EWMA with estimated parameters whereas the fourth specification is a plain GARCH(1,1) which nests the EWMAs within it, since they can be obtained through parameter restrictions. The fifth and final specification is a plain EGARCH(1,1).

Estimates and residual diagnostics of the simple models are contained in tables 5 and 6. The estimate of the Historical specification yield standardised residuals that are uncorrelated according to the AR_{1-10} test. Although this is not the case for the AR_{1-1} test which is not reported, the failure of the AR_{1-10} test to reject the null nevertheless suggests that the historical variance might be difficult to beat out-of-sample. In the RiskMetrics specification the diagnostic tests suggest the values of α and β are suboptimal, since both the standardised residuals and the squared standardised residuals are serially correlated. Indeed, the diagnostic tests of the EWMA supports this picture since there the nulls of uncorrelated and homoscedastic standardised residuals are not rejected. The α, β estimates and diagnostics of the plain GARCH(1,1) specification are almost identical, and the estimate of ω is almost zero. In other words, the two specifications will produce almost identical forecasts. In the EGARCH(1,1) model residuals are also uncorrelated whereas the estimate of the volatility persistence parameter β is high and almost 1 (it is equal to 0.981). The asymmetry parameter γ is not significant at conventional significance levels, thus suggesting the symmetry of the GARCH(1,1) is not so restrictive. Finally, compared with the estimates of ω, α and β in (19) and (20) they are virtually identical here. In other words, adding a mean specification and h_{2t} does not seem to affect the estimates of the variance equation noteworthy.

4 Out-of-sample forecast evaluation

This section proceeds in three steps. The first subsection contains our out-of-sample forecast accuracy comparison, whereas the second contains socalled Mincer-Zarnowitz (1969) regressions of observed volatility on a constant and forecasts. The third and final subsection sheds additional light on the results by examining some of the 1-step forecast trajectories

---

[14]To be more precise, the parameter values are those suggested by the 1995 version of RiskMetrics, which then was part of the merchant bank J.P. Morgan. RiskMetrics is now an independent company and two versions of RiskMetrics have superseded the 1995 May edition, see http://www.riskmetrics.com/techdoc.html. Note also that the parameter values are obtained with a definition of volatility that differs slightly from the one employed here.

more closely.

4.1 Out-of-sample forecast accuracy comparison

Consider a sequence of volatilities $\{V_k\}$ over the forecast periods $k = 1, \ldots, K$ and a corresponding sequence of forecasts $\{\hat{V}_k\}$. Our out-of-sample forecast accuracy measures consist of the mean absolute error (MAE) and the mean squared error (MSE), and are defined as

$$MAE = \frac{1}{K} \sum_{k=1}^{K} |V_k - \hat{V}_k|, \tag{26}$$

$$MSE = \frac{1}{K} \sum_{k=1}^{K} (V_k - \hat{V}_k)^2, \tag{27}$$

respectively.[15] Both measures are symmetric in that they treat positive and negative errors alike, and their only difference is that the latter punishes large errors more severely than the former.[16] Error-based measures are "pure" precision measures in the sense that evaluation is based solely on the discrepancy between the forecast and the actual value. One can make a case for the view that precision-based measures are the most appropriate when evaluating the forecast properties of a certain modelling strategy, since this leaves open what the ultimate use of the model is. On the other hand, this is also a weakness since considerations pertaining to the final use of the model do not enter the evaluation.[17]

The values of the MAE and MSE forecast statistics are contained in tables 7 and 8. In the forecasting literature models with economic covariates are typically championed as producers of accurate long-term forecasts, but not necessarily of short-term forecasts better than those of "naïve" or simple models without economic covariates. Our results seem to contradict this for the short term. On short horizons up to six weeks ahead GETS EMOV1, the specification with actual values on the economic variables of the right hand side, performs well according to both the MAE and MSE statistics. According to the MAE

---

[15]Patton (2005) has recently argued in favour of the MSE and against the MAE in volatility forecast comparison. It should be noted however that his argument applies (under certain assumptions) when the problem to be solved is to choose an $\hat{\sigma}_t^2$ such that expected $L(\sigma_t^2, \hat{\sigma}_t^2)$ is minimised, where $L(\cdot)$ is a loss function. In explanatory modelling on the other hand, as we argued in subsection 2.1, the problem to be solved is to choose an $\hat{\sigma}_t^2$ such that expected $L(r_t^2, \hat{\sigma}_t^2)$ is minimised. This is a qualitatively important difference and it is not clear that Patton's conclusions hold when the problem is formulated in this way.

[16]An alternative to the MAE and MSE are their relative counterparts which scale the error $V_k - \hat{V}_k$ by either V_k or $\hat{V}_k$. The disadvantage with the relative counterparts is that they may favour or disfavour models that either systematically overpredicts or underpredicts, regardless of how well they fare according to the MAE and MSE criteria.

[17]Several other approaches to out-of-sample forecast comparison have been proposed. One consists of adding other ingredients to the evaluation scheme, see for example West et al. (1993) where the expected utility of a risk averse investor serves as the ranking criterion. Similarly, Engle et al. (1993) provide a methodology in which the profitability of a certain trading strategy ranks the forecasts. Yet another approach takes densities as the object of interest, see Diebold et al. (1998), whereas Lopez (2001) has proposed a framework that provides probability forecasts of the event of interest.

it comes 1st, 2nd, 3rd and 3rd up to six weeks ahead, whereas according to the MSE it comes 1st on the same forecast horizons. On longer horizons, however, results are more mixed. For 12 and 18 weeks ahead the GETS EMOV1 comes 10th and 12th (last) according to the MAE, and 9th for both the 12 and 18 horizons according to the MSE. One might suggest that this is due to structural instability, but the results of the other GETS models indicate it is not that straightforward. Indeed, generally the GETS models (EMOV1-4) do notably better according to the MSE measure than the MAE measure, which suggests that the usefulness of GETS modelling increases when larger errors are punished more. The most extreme case is GETS EMOV4 which 12 and 18 weeks ahead comes 12th (last) and 11th (second-to-last) according to the MAE measure, and 4th and 4th according to the MSE measure. This suggests that the poor performance of GETS EMOV3 on longer horizons is not due to parameter instability.

In a practical forecasting situation the actual values on the right hand side of the GETS EMOV1 specification would have to be forecasted, and GETS EMOV2 and GETS EMOV3 try to mimic such a situation. Both models are relatively consistent and perform comparatively well on shorter horizons, in particular according to the MSE measure, but as long as no statistical tests are involved it is unclear whether it is significantly superior or inferior to any particular model. On the first four horizons the GETS EMOV2 comes 3rd, 5th, 5th and 6th according to the MAE, and 4th, 4th, 4th and 5th according to the MSE. On the same horizons the GETS EMOV3 comes 4th, 6th, 7th and 8th according to the MAE, and 3rd, 3rd, 3th and 4rd according to the MSE. For 12 and 18 weeks ahead the GETS EMOV4 and GETS EMOV5 do not fare well according to the MAE, since they come 7th or worse. According to the MSE on the other hand they come 4th or better.

Although the MAE and MSE measures suggest that the GETS models perform relatively well compared with the other models, it should be stressed that so do some of the simple models at times. For example, according to the MAE the Historical specification, that is, the constant model of volatility, comes 2nd at the 1 week horizon, 4rd at the 2 and 3 weeks horizons, and according to the MSE it is distinguishable from the best model 12 and 18 weeks ahead only at the third decimal. Similarly, the Range EMOV6, that is, the specification that uses lagged range volatility as predictor, comes 1st on the 2, 3 and 6 week horizons according to the MAE. The RiskMetrics, GARCH and EGARCH specifications do not do particularly well at short horizons, that is, on horizons in which one would expect them to do well. Not once does any of the five specifications beat Historical 1 to 3 weeks ahead according to both MAE and MSE.

4.2   1-step Mincer-Zarnowitz regressions

A simple statistical way of evaluating forecast models is by regressing the variable to be forecasted on a constant and on the forecasts, socalled Minzer-Zarnowitz (1969) regressions, see Andersen and Bollerslev (1998) and Patton (2005) for a discussion of their use in

volatility forecast evaluation.[18] In our case this proceeds by estimating the specification

$$r_t^2 = a + b\hat{V}_t + e_t, \tag{28}$$

where r_t^2 is observed volatility, $\hat{V}_t$ is the 1-step forecast and e_t is the error term. Ideally, a should equal zero and b should equal one—since these constitute conditions for "unbiasedness", and the fit should be high. Table 9 contains the regression output.[19] One specification stands out according to the majority of the criteria, namely GETS EMOV1. Its estimate of a is not significantly different from zero, the estimate of b is positive and significantly different from zero, the joint restriction $a = 0, b = 1$ is not rejected at conventional significance levels, and its R^2 is 0.26. This is substantially higher than any of the R^2s cited in Andersen and Bollerslev (1998, pp. 890-891) (the typical R^2 they cite is around 0.03 and the highest is 0.11), and must be very close to—if not exceeding—their population upper bound of R^2:

> "..with conditional Gaussian errors the $R_{(m)}^2$ from a correctly specified GARCH(1,1) model is bounded from above by $\frac{1}{3}$, while with conditional fat-tailed errors the upper bound is even lower. Moreover, with realistic parameter values for $\alpha_{(m)}$ and $\beta_{(m)}$, the population value for the $R_{(m)}^2$ statistic is significantly below this upper bound"—Andersen and Bollerslev (1998, p. 892).

In other words, although the unusually high R^2 of GETS EMOV3 might be due to sample specificity, it nevertheless suggests the poor forecasting performance of r_t^2 by ARCH-models can be improved upon substantially. Moreover, apart from the RiskMetrics specification Historical beats the other five members of the ARCH-family (EWMA, GARCH(1,1), EGARCH(1,1), GARCH(1,1)+ and EGARCH(1,1)+), and the four models GETS EMOV1-4 perform better than Historical according to R^2. Also, in none of these four specifications is neither a significantly different from zero, nor is the joint restriction $a = 0, b = 1$ rejected. Apart from Historical and RiskMetrics, the restriction $a = 0, b = 1$ is rejected at the 5% level in all the ARCH-specifications, and a is significantly different from zero.

4.3   Explaining the forecast results

An important part of an out-of-sample study consists of explaining the results, and to this end figure 6 provides a large part of the answer. The figure contains the out-of-sample trajectories of squared NOK/EUR log-returns in percent r_t^2, the 1-step forecasts of GETS EMOV1 and the 1-step forecasts of Historical, and the figure provides some interesting

---

[18]It should be stressed though that their discussion concerns forecasting the *latent* volatility rather than observed volatility, as is the case in this study.

[19]Patton (2005, footnote on p. 6) has noted that the residuals in Mincer-Zarnowitz regressions typically are serially correlated and that this should be taken into account by using (say) Newey and West (1987) standard errors. In our case the residuals are not serially correlated according to standard tests, but admittedly it might be undetectable due to our relatively small sample.

insights on the forecast accuracy results. First, the series of r_t^2 seems to be characterised by some occasional large values but little volatility persistence in the sense that large values do not tend to follow each other. Indeed, only at two instances is a large value followed by another, and for a relatively large portion of the sample r_t^2 stays rather low. This explains to some extent the forecast accuracy of Historical. It also explains to some extent the success of Range EMOV6, since its trajectory—which is not contained in the figure—is very similar to Historical's. Second, in the 5th and in the 11th weeks of the forecast sample Norges Bank changed its main policy interest rate. This is reflected in the large values of r_t^2 in the 5th and 11th weeks, and explains the forecast accuracy of GETS EMOV1 (it contains policy interest rate changes as a predictor) and its unusually high R-squared in the 1-step forecast regressions. Finally, the other explanatory variables included in GETS EMOV1 are probably the reason why it also follows r_t^2 relatively well at other instances when r_t^2 moves substantially. But it should be noted that the explanatory regressors in GETS EMOV1 also seem to induce notable forecast error on at least two occasions. In other words, although the forecast success of GETS EMOV1 is due to the explanatory variables, there are also signs that they may have the opposite effect, namely increasing the forecast error. All in all, then, the forecast results suggest the GETS EMOV1 is useful for conditional forecasting but that it does not improve upon the forecasting by simple models when the explanatory variables are unchanging or move little.

5   Conclusions

This study has evaluated the out-of-sample forecast accuracy of GETS derived models of weekly NOK/EUR volatility. The GETS specification that uses actual values of uncertain information is found to perform particularly well when it is able to explain big movements in the exchange rate, but not necessarily better than simple models like the constant volatility model when the exchange rate does not move much or when it is unable to explain the movement. Models of the GARCH(1,1) and EGARCH(1,1) types do not fare particularly well and the reason is that large values (in the absolute sense) of returns do not seem to come in pairs nor in longer sequences at the weekly frequency. Rather, big movements in the exchange rate seems more to be a "one off" phenomenon.

Overall, then, our results suggest GETS derived models of observable volatility are not inferior to the comparison models, and that they are particularly useful in conditional forecasting. This suggests several lines for further research. First, the generality of our results must be established. Is GETS-modelling of financial volatility useful on higher frequencies than the weekly? On other financial assets? Second, contrary to McAleer's (2005) assertion, automated GETS-modelling of financial volatility can be readily implemented and should be investigated more fully.

18

References

Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*, 885–905.

Andersen, T. G., T. Bollerslev, and S. Lange (1999). Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon. *Journal of Empirical Finance 6*, 457–477.

Bauwens, L., D. Rime, and G. Sucarrat (2006). Exchange Rate Volatility and the Mixture of Distribution Hypothesis. *Empirical Economics 30*, 889–911.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics 31*, 307–327.

Bollerslev, T. and J. Wooldridge (1992). Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances. *Econometric Reviews 11*, 143–172.

Campos, J., N. R. Ericsson, and D. F. Hendry (2005). General-to-Specific Modeling: An Overview and Selected Bibliography. In J. Campos, D. F. Hendry, and N. R. Ericsson (Eds.), *General-to-Specific Modeling*. Cheltenham: Edward Elgar Publishing.

Diebold, F., T. Gunther, and A. Tay (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review 39*, 863–883.

Engle, R., C.-H. Hong, A. Kane, and J. Noh (1993). Arbitrage Valuation of Variance Forecasts with Simulated Options. In D. Chance and R. Trippi (Eds.), *Advances in Futures and Options Research*. Greenwich, Connecticut: JAI Press.

Geweke, J. (1986). Modelling the Persistence of Conditional Variance: A Comment. *Econometric Reviews 5*, 57–61.

Gilbert, C. L. (1989). Lse and the british approach to time series econometrics. *Oxford Economic Papers 41*, 108–128.

Gilbert, C. L. (1990). Professor Hendry's Econometric Methodology. In C. W. Granger (Ed.), *Modelling Economic Series*. Oxford: Oxford University Press. Earlier publised in Oxford Bulletin of Economics and Statistics 48 (1986), pp. 283-307.

Hendry, D. F. (1987). Econometric Methodology: A Personal Perspective. In T. F. Bewley (Ed.), *Fifth World Congress of the Econometric Society*. Oxford: Oxford University Press.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (1997). On Congruent Econometric Relations: A Comment. *Carnegie-Rochester Conference Series on Public Policy 47*, 163–190.

Hendry, D. F. (2003). J. Denis Sargan and the Origins of LSE Econometric Methodology. *Econometric Theory 19*, 457–480.

Hendry, D. F. and H.-M. Krolzig (1999). Improving on 'Data Mining Reconsidered' by K.D. Hoover and S.J. Perez. *Econometrics Journal 2*, 202–219.

Hendry, D. F. and H.-M. Krolzig (2001). *Automatic Econometric Model Selection using PcGets*. London: Timberlake Consultants Press.

Hendry, D. F. and H.-M. Krolzig (2004). We Ran One Regression. Research in Progress. Http://www.nuff.ox.ac.uk/Economics/Papers/2004/w17/OneReg.pdf.

Hendry, D. F. and J.-F. Richard (1990). On the Formulation of Empirical Models in Dynamic Econometrics. In C. W. Granger (Ed.), *Modelling Economic Series*. Oxford: Oxford University Press. Earlier in Journal of Econometrics 20 (1982), pp. 3-33.

Hoover, K. D. and S. J. Perez (1999). Data Mining Reconsidered: Encompassing and the General-to-Specific Approach to Specification Search. *Econometrics Journal 2*, 167–191.

Hull, J. C. (2000). *Options, Futures and Other Derivatives*. London: Prentice Hall International. Fourth Edition.

Jansen, E. (2002). Statistical Issues in Macroeconomic Modelling. *Scandinavian Journal of Statistics 29*, pp. 193–217.

Jarque, C. and A. Bera (1980). Efficient Tests for Normality, Homoskedasticity, and Serial Independence of Regression Residuals. *Economics Letters 6*, 255–259.

Jorion, P. (1995). Predicting Volatility in the Foreign Exchange Markets. *The Journal of Finance 50*, 507–528.

Ljung, G. and G. Box (1979). On a Measure of Lack of Fit in Time Series Models. *Biometrika 66*, 265–270.

Lopez, J. A. (2001). Evaluating the Predictive Accuracy of Volatility Models. *Journal of Forecasting 20*, 87–109.

Lovell, M. C. (1983). Data Mining. *The Review of Economics and Statistics 65*, 1–12.

McAleer, M. (2005). Automated inference and learning in modeling financial volatility. *Econometric Theory 21*, 232–261.

Mincer, J. and V. Zarnowitz (1969). The Evaluation of Economic Forecasts. In J. Zarnowitz (Ed.), *Economic Forecasts and Expectations*. New York: National Bureau of Economic Research.

Mizon, G. (1995). Progressive Modeling of Macroeconomic Time Series: The LSE Methodology. In K. D. Hoover (Ed.), *Macroeconometrics. Developments, Tensions and Prospects*. Kluwer Academic Publishers.

Nelson, D. B. (1991). Conditional Heteroscedasticity in Asset Returns: A New Approach. *Econometrica 51*, 485–505.

Newey, W. and K. West (1987). A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*, 703–708.

Pantula, S. (1986). Modelling the Persistence of Conditional Variance: A Comment. *Econometric Reviews 5*, 71–73.

Patton, A. J. (2005). Volatility Forecast Evaluation and Comparison Using Imperfect Volatility Proxies. Downloadable from http://fmg.lse.ac.uk/~patton/research.html.

West, K., H. Edison, and D. Cho (1993). A Utility Based Comparison of Some Models for Exchange Rate Volatility. *Journal of International Economics 35*, 23–45.

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity. *Econometrica 48*, 817–838.

White, H. (1990). A consistent model selection procedure based on emphm-testing. In C. W. Granger (Ed.), *Modelling Economic Series: Readings in Econometric Methodology*. Oxford: Clarendo Press.

Appendix: Data transformations and sources

The data transformations were undertaken in Ox 3.4 and EViews 5.1.

S_t BID NOK/1EUR closing value of the last trading day of week t. Before 1.1.1999 the BID NOK/1EUR rate is obtained by the formula BID NOK/100DEM $\times$ 0.0195583, where 0.0195583 is the official DEM/1EUR conversion rate 1.95583 DEM = 1 EUR divided by 100. The source of the BID NOK/100DEM series is Olsen Financial Technologies and the source of the BID NOK/1EUR series is Reuters.

r_t $(\log S_t - \log S_{t-1}) \times 100$

V_t^w $\{\{\log[S_t + I(S_t = S_{t-1}) \times 0.0009] - \log(S_{t-1})\} \times 100\}^2$. $I(S_t = S_{t-1})$ is an indicator function equal to 1 if $S_t = S_{t-1}$ and 0 otherwise, and $S_t = S_{t-1}$ occurs for $t = 10/6/1994$, $t = 19/8/1994$ and $t = 17/2/2000$.

v_t^w $\log V_t^w$

V_t^r $\sum_n [\log(S_n/S_{n-1}) \times 100]^2$, where $n = 1(t), 2(t), ..., N(t)$ and $1(t)-1 = N(t-1)$. $S_{1(t)}$ is the first BID NOK/1EUR opening exchange rate of week t, $S_{2(t)}$ is the first closing rate, $S_{3(t)}$ is the second opening rate, and so on, with $S_{N(t)}$ denoting the last closing rate of week t, that is, $S_{N(t)} = S_t$.

v_t^r $\log V_t^r$

V_t^{hl} $[\log(S_t^h/S_t^l) \times 100]^2$, where S_t^h and S_t^l are the maximum and minimum values of bid NOK/EUR in week t.

v_t^{hl} $\log V_t^{hl}$

M_t BID USD/EUR closing value of the last trading day of week t. Before 1.1.1999 the BID USD/EUR rate is obtained with the formula 1.95583/(BID DEM/USD). The source of the BID DEM/USD and BID USD/EUR series is Reuters.

m_t $\log M_t$

M_t^w $\{\{\log[M_t + I(M_t = M_{t-1}) \times 0.0009] - \log(M_{t-1})\} \times 100\}^2$. $I(M_t = M_{t-1})$ is an indicator function equal to 1 if $M_t = M_{t-1}$ and 0 otherwise.

m_t^w $\log M_t^w$

Q_t Weekly number of NOK/EUR quotes (NOK/100DEM before 1.1.1999). The underlying data is a daily series from Olsen Financial Technologies, and the weekly values are obtained by summing the values of the week.

q_t $\log Q_t$. This series is "synthetic" in that it has been adjusted for changes in the underlying quote collection methodology at Olsen Financial Technologies. More precisely q_t has been generated under the assumption that Δq_t is equal to zero in the weeks containing Friday 17 August 2001 and Friday 5 September 2003, respectively. In the first week the underlying feed was changed from Reuters to Tenfore, and on the second a feed from Oanda was added.

Δq_t $q_t - q_{t-1}$. The values of this series has been set to zero in the weeks containing Friday 24 August 2001 and Friday 5 September 2003, respectively.

O_t Closing value of the Brent Blend spot oilprice in USD per barrel in the last trading day of week t. The untransformed series is Bank of Norway database series D2001712.

o_t $\log O_t$

O_t^w $\{\{\log[O_t + I(O_t = O_{t-1}) \times 0.009] - \log(O_{t-1})\} \times 100\}^2$. $I(O_t = O_{t-1})$ is an indicator function equal to 1 if $O_t = O_{t-1}$ and 0 otherwise, and $O_t = O_{t-1}$ occurs three times, for $t = 1/7/1994$, $t = 13/10/1995$ and $t = 25/7/1997$.

o_t^w $\log O_t^w$

X_t Closing value of the main index of the Norwegian Stock Exchange (TOTX) in the last trading day of week t. The source of the daily untransformed series is EcoWin series ew:nor15565.

x_t $\log X_t$

X_t^w $\{[\log(X_t/X_{t-1})] \times 100\}^2$. $X_t = X_{t-1}$ does not occur for this series.

x_t^w $\log X_t^w$

U_t Closing value of the composite index of the New York Stock Exchange (the NYSE index) in the last trading day of week t. The source of the daily untransformed series is EcoWin series ew:usa15540.

U_t^w $\{[\log(U_t/U_{t-1})] \times 100\}^2$. $U_t = U_{t-1}$ does not occur for this series.

u_t^w $\log U_t^w$

IR_t^{emu} Average of closing values of the 3-month market interest rates of the European Monetary Union (EMU) countries in the last trading day of week t. The source of the daily untransformed series is EcoWin series ew:emu36103.

ir_t^{emu} $(\Delta IR_t^{emu})^2$.

F_t The Norwegian central bank's main policy interest-rate, the socalled "folio", at the end of the last trading day of week t. The source of the untransformed daily series is Norges Bank's webpages.

f_t^a $|\Delta F_t| \times I_a$, where I_a is an indicator function equal to 1 in the period 1 January 1999 - Friday 30 March 2001 and 0 elsewhere

f_t^b $|\Delta F_t| \times I_b$, where I_b is an indicator function equal to 1 after Friday 30 March 2001 and 0 before

id_t Russian moratorium impulse dummy, equal to 1 in the week containing Friday 28 August 1998 and 0 elsewhere.

sd_t Step dummy, equal to 0 before 1997 and 1 thereafter.

ia_t Skewness term, equal to 1 when $r_t > 0$ and 0 otherwise.

h_{lt} $l = 1, 2, \ldots, 8$. Holiday variables with values equal to the number of official Norwegian holidays that fall on weekdays. For example, if 1 January falls on a Saturday then h_{1t} is equal to 0, whereas if 1 January falls on a Monday, then h_{1t} is equal to 1. h_{2t} is associated with Maundy Thursday and Good Friday and thus always equal to 2, h_{3t} with Easter Monday and thus always equal to 1, h_{4t} with Labour Day (1 May), h_{5t} with the Norwegian national day (17 May), h_{6t} with Ascension Day, h_{7t} with Whit Monday and h_{8t} with Christmas (Christmas Day and Boxing Day). Source: Http://www.timeanddate.com.

Table 1: GUM and GETS regressions of log of weekly NOK/EUR volatility on both certain and uncertain information

Parameter	(11)		(12)	
	Est.	Pval.	Est.	Pval.
b_0	-3.304	0.03	-3.035	0.00
b_1	0.013	0.76		
b_2	0.070	0.07	0.079	0.00
b_3	0.093	0.03		
b_4	-0.001	0.99		
b_5	0.063	0.76		
b_6	1.024	0.00	1.066	0.00
b_7	0.067	0.13		
b_8	0.021	0.65		
b_9	0.125	0.01	0.119	0.00
b_{10}	0.113	0.01		
b_{11}	-0.256	0.83		
b_{12}	3.775	0.00	3.751	0.00
b_{13}	4.797	0.02	4.819	0.00
b_{14}	1.130	0.00	1.238	0.00
b_{15}	-0.127	0.52		
b_{16}	-0.025	0.85		
b_{17}	-1.207	0.16		
b_{18}	-0.141	0.62		
b_{19}	0.330	0.64		
b_{20}	-0.710	0.22		
b_{21}	0.195	0.71		
b_{22}	0.653	0.25		
b_{23}	0.019	0.98		
b_{24}	-0.036	0.96		
R^2	0.21		0.20	
AR_{1-10}	5.11	0.88	3.07	0.98
$ARCH_{1-10}$	7.00	0.73	8.71	0.56
$Het.$	39.72	0.44	11.46	0.41
$Hetero.$	179.98	0.92	14.60	0.95
$Obs.$	568		569	

Note: See table 2.

Table 2: GUM and GETS Regressions of log of weekly NOK/EUR volatility on certain information only

Parameter	(15)		(16)	
	Est.	Pval.	Est.	Pval.
b_0	-2.918	0.00	-2.946	0.00
b_1	0.030	0.48		
b_2	0.077	0.04	0.088	0.00
b_3	0.093	0.03		
b_4	-0.020	0.64		
b_{13}	1.448	0.00	1.428	0.00
b_{14}	-0.009	0.96		
b_{16}	-0.024	0.86		
b_{17}	-0.820	0.30		
b_{18}	-0.500	0.07	-0.478	0.08
b_{19}	0.435	0.59		
b_{20}	-0.751	0.17		
b_{21}	0.170	0.72		
b_{22}	0.346	0.55		
b_{23}	-0.446	0.56		
b_{24}	-0.625	0.37		
R^2	0.14		0.13	
AR_{1-10}	9.14	0.52	4.22	0.94
$ARCH_{1-10}$	8.59	0.57	9.25	0.51
$Het.$	26.96	0.17	6.39	0.17
$Hetero.$	90.65	0.66	7.43	0.39
$Obs.$	568		569	

Note: Computations are in EViews 5.1 with OLS estimation. All specifications use heteroscedasticity consistent standard errors of the White (1980) type, *Pval* stands for *p*-value and corresponds to a two-sided test with zero as null, AR_{1-10} is the χ^2 version of the Lagrange-multiplier test for serially correlated residuals up to lag 10, $ARCH_{1-10}$ is the χ^2 version of the Lagrange-multiplier test for serially correlated squared residuals up to lag 10, and *Het.* and *Hetero.* are the χ^2 versions of White's (1980) heteroscedasticity tests without and with cross products, respectively.

Table 3: Regressions of log of weekly volatility on log of lagged realised volatility and on log of lagged range volatility

Parameter	(17)		(18)	
	Est.	Pval.	Est.	Pval.
b_0	-2.774	0.00	-3.018	0.00
b_1	0.304	0.00	0.348	0.00
b_{14}	1.158	0.00	1.057	0.00
b_{18}	-0.510	0.07	-0.527	0.06
R^2	0.13		0.13	
AR_{1-10}	9.35	0.50	10.06	0.44
$ARCH_{1-10}$	11.20	0.34	10.46	0.40
$Het.$	5.91	0.21	3.21	0.52
$Hetero.$	7.64	0.37	4.51	0.72
$Obs.$	571		572	

Note: See table 2.

Table 4: ARCH models of r_t with certain information.

Parameter	(19)		(20)	
	Est.	Pval.	Est.	Pval.
b_0	-0.017	0.42	-0.007	0.77
b_1	-0.055	0.33	-0.067	0.21
ω	0.009	0.07	-0.209	0.00
α	0.129	0.01	0.285	0.00
β	0.877	0.00	0.983	0.00
γ_0			0.012	0.86
γ_1	-0.043	0.08	-0.171	0.35
$LogL.$	-658.74		-571.98	
AR_{1-10}	9.64	0.47	14.62	0.15
$ARCH_{1-10}$	4.80	0.90	5.75	0.84
JB	664.93	0.00	417.11	0.00
$Obs.$	571		571	

Note: Computations are in EViews 5.1 with robust standard errors of the Bollerslev and Wooldridge (1992) type. *Pval* stands for *p*-value and corresponds to a two-sided test with zero as null, *LogL* stands for log-likelihood, AR_{1-10} is the Ljung and Box (1979) test for serial correlation in the standardised residuals up to lag 10, $ARCH_{1-10}$ is the Ljung and Box (1979) test for serial correlation in the squared standardised residuals up to lag 10, and *JB* is the Jarque and Bera (1980) test for non-normality in the standardised residuals.

Table 5: Historical, RiskMetrics and EWMA models of r_t.

Parameter	(21)		(22)		(23)	
	Est.	Pval.	Est.	Pval.	Est.	Pval.
ω	0.681	0.00				
α			0.060		0.159	0.01
β			0.940		0.877	0.00
LogL.	-320.29				-583.63	
AR_{1-10}	8.94	0.54	31.49	0.00	12.88	0.23
$ARCH_{1-10}$	9.49	0.49	46.38	0.00	5.32	0.72
JB	87.85	0.00	145K	0.00	565.09	0.00
Obs.	261		572		572	

Note: Computations are in G@RCH 4.0 and EViews 5.1 with robust standard errors of the Bollerslev and Wooldridge (1992) type. Otherwise see table 4.

Table 6: Plain ARCH models of r_t.

Parameter	(24)		(25)	
	Est.	Pval.	Est.	Pval.
ω	0.006	0.14	-0.223	0.00
α	0.146	0.01	0.292	0.00
β	0.867	0.00	0.981	0.00
γ			0.017	0.79
LogL.	-580.96		-576.75	
AR_{1-10}	11.43	0.32	12.27	0.27
$ARCH_{1-10}$	4.97	0.89	6.45	0.78
JB	635.55	0.00	377.88	0.00
Obs.	572		572	

Note: See table 4.

Table 7: MAE forecast statistics

	1-step	2-step	3-step	6-step	12-step	18-step
GETS EMOV1	**0.74**	0.75	0.65	0.62	0.66	0.74
GETS EMOV2	0.83	0.83	0.72	0.67	0.64	0.66
GETS EMOV3	0.85	0.85	0.73	0.70	0.66	0.68
GETS EMOV4	0.90	0.90	0.77	0.74	0.70	0.73
Realised EMOV5	0.90	0.78	0.63	0.62	0.58	0.61
Range EMOV6	0.96	**0.74**	**0.60**	**0.60**	0.56	0.59
GARCH(1,1)+	0.98	0.91	0.72	0.65	0.56	0.62
EGARCH(1,1)+	1.02	1.96	1.48	0.85	**0.55**	0.65
Historical	0.81	0.82	0.70	0.69	0.65	0.67
RiskMetrics	0.92	1.00	0.79	0.77	0.64	**0.58**
EWMA	1.06	1.24	0.97	0.83	0.63	0.61
GARCH(1,1)	0.99	1.16	0.89	0.78	0.61	0.61
EGARCH(1,1)	1.00	0.88	0.75	0.66	0.59	0.67

Note: Bold value indicates minimum in its column.

Table 8: MSE forecast statistics

	1-step	2-step	3-step	6-step	12-step	18-step
GETS EMOV1	**1.61**	**1.64**	**1.07**	**1.07**	1.25	1.40
GETS EMOV2	2.07	2.10	1.21	1.20	1.02	1.12
GETS EMOV3	2.06	2.09	1.21	1.19	1.02	**1.12**
GETS EMOV4	2.04	2.06	1.22	1.19	1.03	1.12
Realised EMOV5	2.14	2.23	1.24	1.25	1.06	1.17
Range EMOV6	2.25	2.37	1.39	1.41	1.21	1.34
GARCH(1,1)+	2.31	2.23	1.25	1.24	1.20	1.42
EGARCH(1,1)+	2.29	4.82	2.81	1.35	1.20	1.50
Historical	2.13	2.16	1.20	1.19	**1.02**	1.12
RiskMetrics	2.18	2.66	1.37	1.47	1.17	1.21
EWMA	2.44	4.76	2.54	2.05	1.31	1.39
GARCH(1,1)	2.33	4.25	2.18	1.85	1.28	1.39
EGARCH(1,1)	2.29	2.19	1.48	1.43	1.33	1.56

Note: Bold value indicates minimum in its column.

Table 9: Mincer-Zarnowitz regressions of r_t^2 on a constant and 1-step out-of-sample forecasts ($K = 61$)

	a	b	R^2	$Pval.$
GETS EMOV1	-0.17	1.36	0.26	0.79
	[0.63]	[0.02]		
GETS EMOV2	-0.32	1.65	0.03	0.76
	[0.78]	[0.36]		
GETS EMOV3	-0.32	1.54	0.03	0.90
	[0.78]	[0.36]		
GETS EMOV4	-0.20	1.20	0.03	0.91
	[0.82]	[0.31]		
Realised EMOV5	0.75	0.10	0.00	0.64
	[0.37]	[0.91]		
Range EMOV6	1.42	-0.66	0.01	0.00
	[0.03]	[0.26]		
GARCH(1,1)+	0.92	-0.08	0.00	0.00
	[0.01]	[0.80]		
EGARCH(1,1)+	0.93	-0.09	0.00	0.02
	[0.02]	[0.83]		
Historical	-	1.24	0.00	0.39
		[0.00]		
RiskMetrics	0.93	-0.10	0.00	0.13
	[0.08]	[0.86]		
EWMA	0.95	-0.10	0.00	0.00
	[0.01]	[0.70]		
GARCH(1,1)	0.96	-0.12	0.00	0.00
	[0.01]	[0.69]		
EGARCH(1,1)	0.98	-0.14	0.00	0.01
	[0.01]	[0.71]		

Note: Numbers in square brackets denote the p-values of a two-sided coefficient hypothesis test with zero as the null hypothesis, and the last column denotes the p-value of a $\chi^2(2)$ Wald test of the joint restriction $a = 0, b = 1$. Otherwise see table 2.
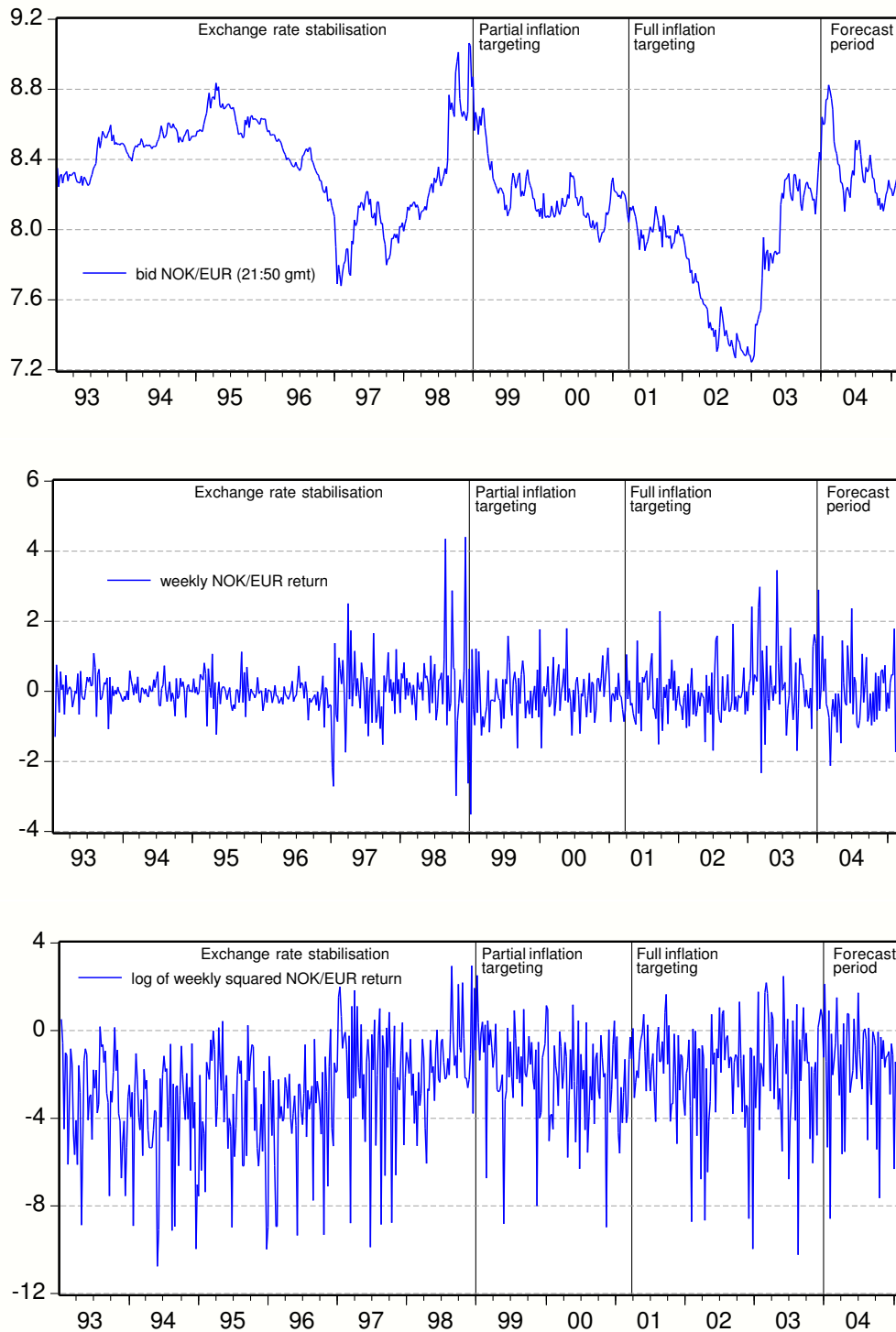
Figure 1: Bid NOK/EUR at 21:50 GMT in the last trading day of the week (denoted S_t in the text) in the upper graph, log-return r_t in the middle graph and log of r_t^2 in the bottom graph from 8 January 1993 to 25 February 2005.
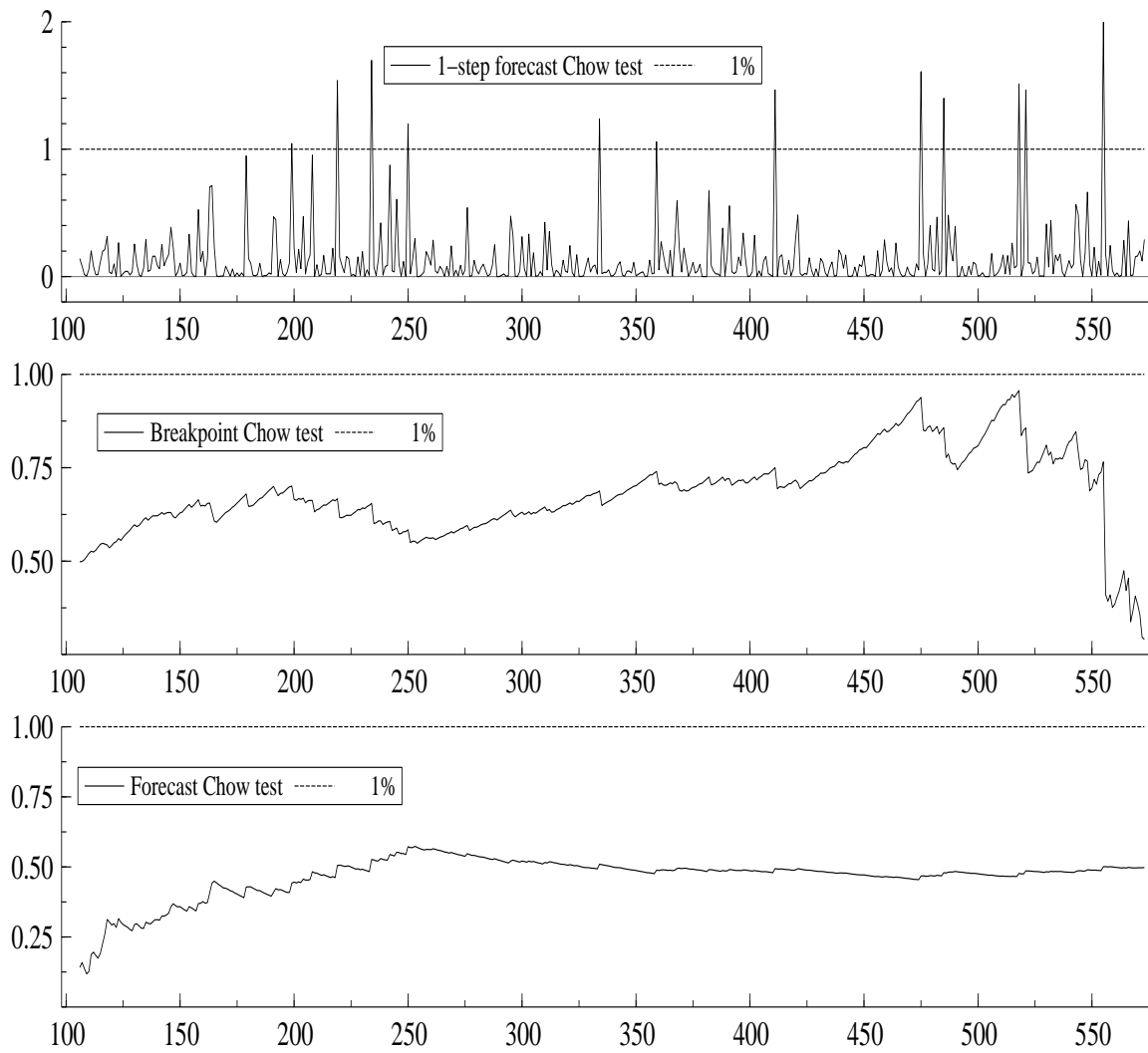
Figure 2: Recursive analysis of GUM EMOV1. Computations are in PcGive 10.4 with OLS and initialisation at observation number 100.
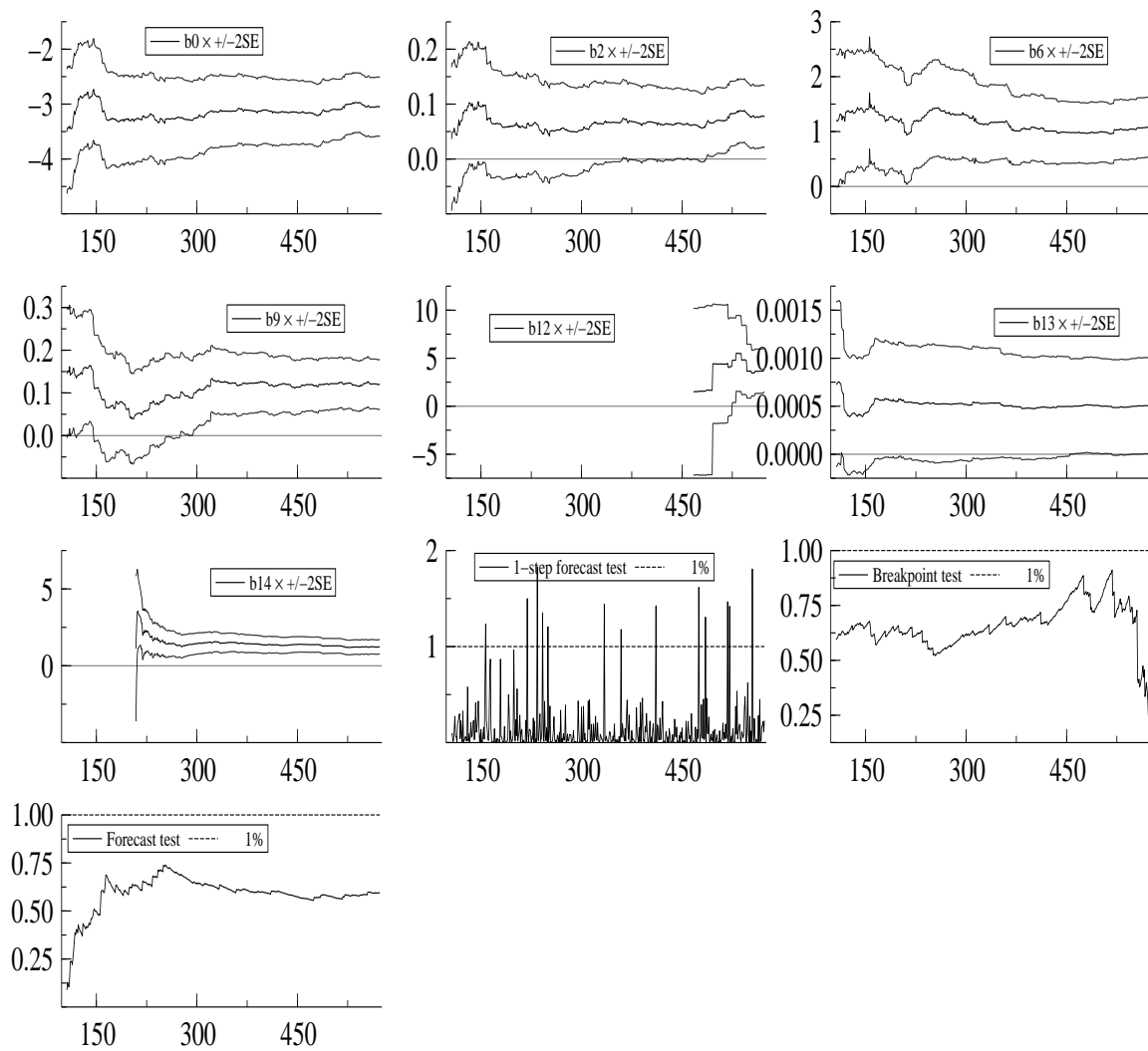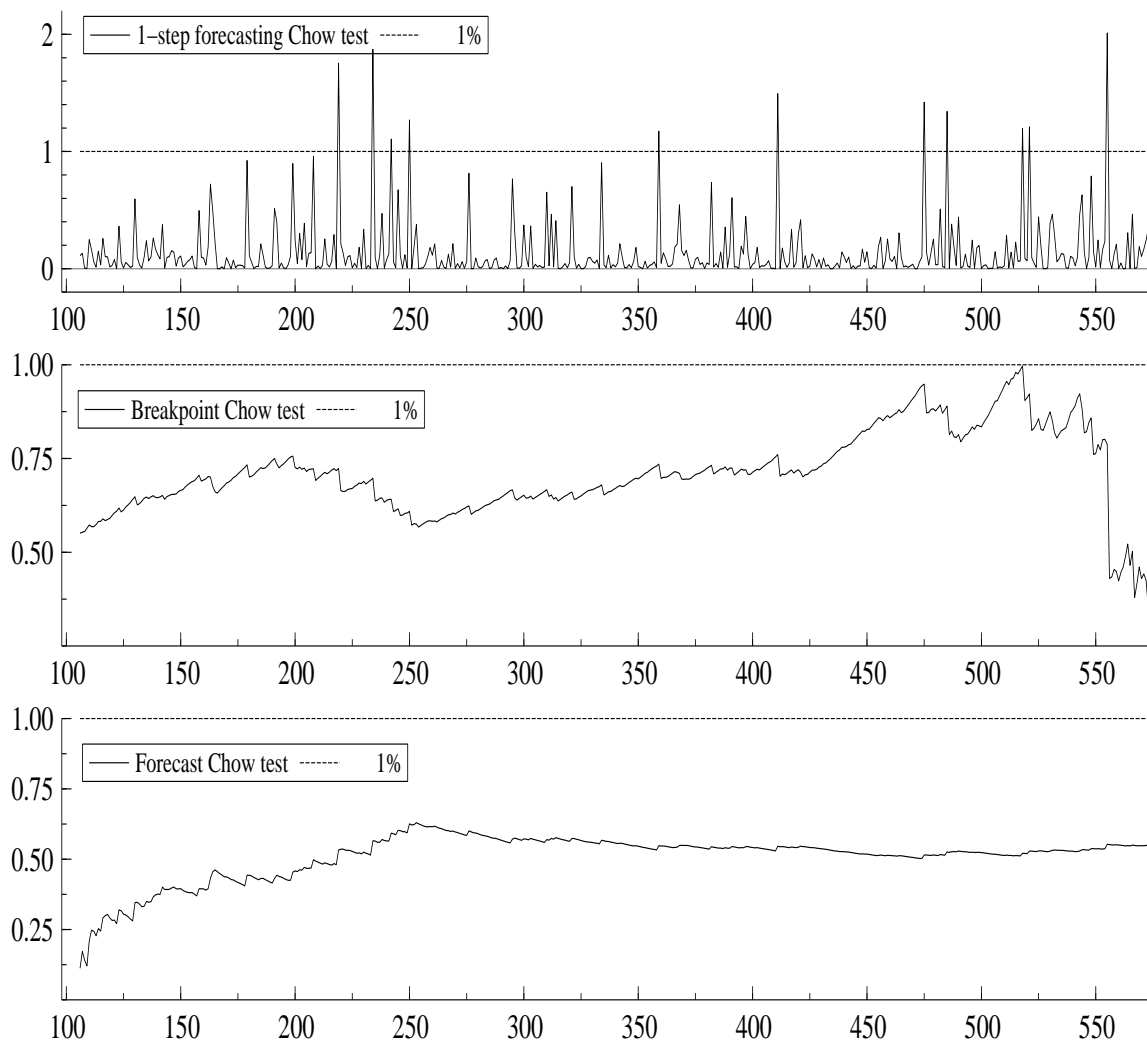
Figure 3: Recursive analysis of GETS EMOV1. Computations are in PcGive 10.4 with OLS and initialisation at observation number 100.

Figure 4: Recursive analysis of GUM EMOV4. Computations are in PcGive 10.4 with OLS and initialisation at observation number 100.
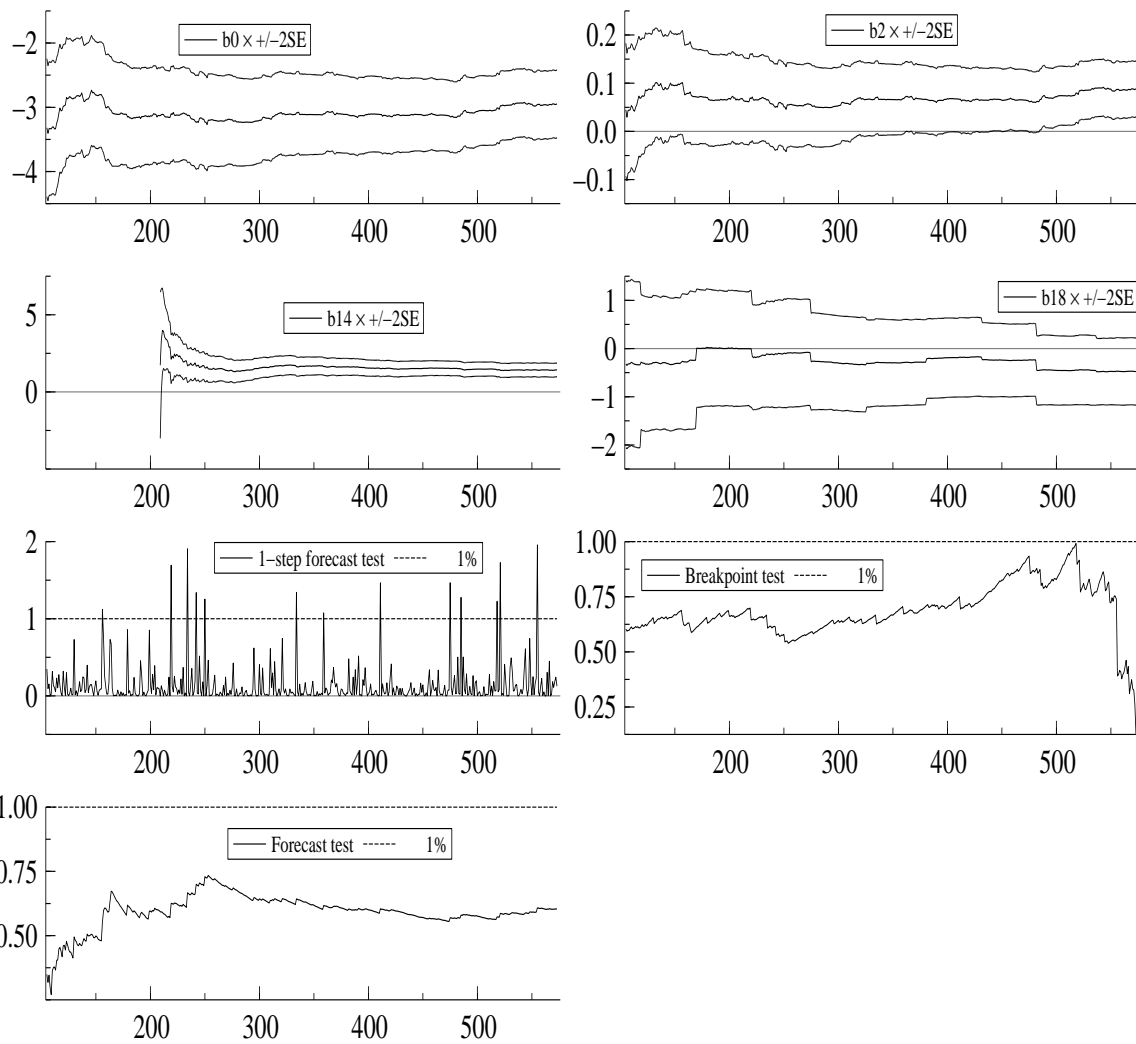
Figure 5: Recursive analysis of GETS EMOV4. Computations are in PcGive 10.4 with OLS and initialisation at observation number 100.
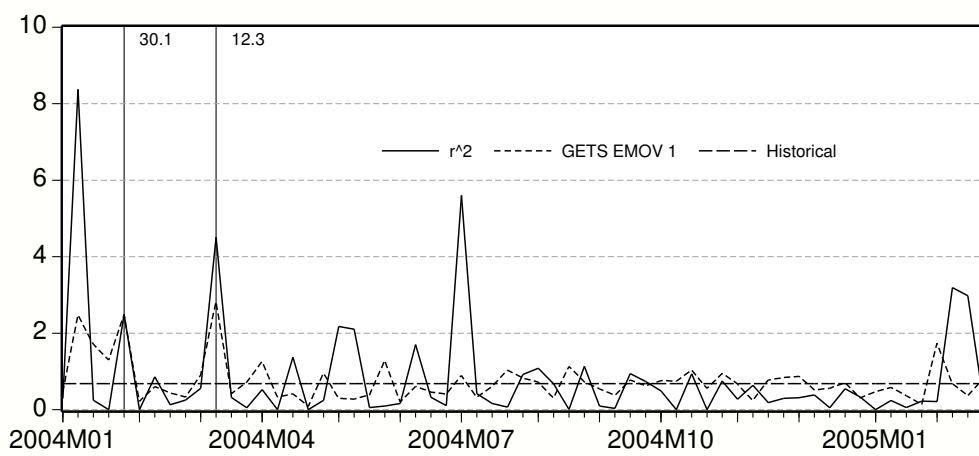
Figure 6: Out-of-sample trajectories of r_t^2, GETS EMOV1 and Historical. Vertical lines indicate weeks in which Norges Bank changed their main policy interest rate.